UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/033,199	12/28/2001	Shivnath Babu	BABU 1-10-42	8231

47394   7590   09/19/2005
HITT GAINES, PC
LUCENT TECHNOLOGIES INC.
PO BOX 832570
RICHARDSON, TX 75083

EXAMINER
LERNER, MARTIN

ART UNIT	PAPER NUMBER
2654	

DATE MAILED: 09/19/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

	Application No.	Applicant(s)
Office Action Summary	10/033,199	BABU ET AL.
	Examiner	Art Unit	
	Martin Lerner	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on <u>25 August 2005</u>.

2a) ☐ This action is **FINAL.**   2b) ☒ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) *1 to 24* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) *1 to 22 and 24* is/are rejected.

7) ☒ Claim(s) *23* is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☒ The drawing(s) filed on <u>28 December 2001</u> is/are: a) ☐ accepted or b) ☒ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All   b) ☐ Some * c) ☐ None of:

      1. ☐ Certified copies of the priority documents have been received.

      2. ☐ Certified copies of the priority documents have been received in Application No. _____.

      3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Drawings

1.   The drawings are objected to because Figures 6 and 7 contain handwritten material and are too small to easily read.

2.   Corrected drawing sheets in compliance with 37 CFR 1.121(d) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The figure or figure number of an amended drawing should not be labeled as "amended." If a drawing figure is to be canceled, the appropriate figure must be removed from the replacement sheet, and where necessary, the remaining figures must be renumbered and appropriate changes made to the brief description of the several views of the drawings for consistency. Additional replacement sheets may be necessary to show the renumbering of the remaining figures. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the examiner does not accept the changes, the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

Claim Rejections - 35 USC § 102

3.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that

form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by
> another filed in the United States before the invention by the applicant for patent or (2) a patent
> granted on an application for patent by another filed in the United States before the invention by the
> applicant for patent, except that an international application filed under the treaty defined in section
> 351(a) shall have the effects for purposes of this subsection of an application filed in the United States
> only if the international application designated the United States and was published under Article 21(2)
> of such treaty in the English language.

4.   Claims 1, 9, and 17 are rejected under 35 U.S.C. 102(e) as being anticipated by

Fayyad et al. ('882).

        Fayyad et al. ('882) discloses a system and method for database management,

comprising:

        "a table modeler that discovers at least one model of data mining models with

guaranteed error bounds of at least one attribute in said data table in terms of other

attributes in different columns of said data table" – the invention evaluates a database

10 having many records stored on storage devices; each record in the database 10 has

many attributes or fields which for a representative database might include age, income,

number of years of employment, census data, etc. (column 4, line 60 to column 5, line

2); implicitly, a plurality of records, where each record has a number of attributes, is a

table; a data clustering model ("table modeler") is produced that implements a data

mining engine for answering queries about data records in the database (column 5,

lines 20 to 25); accuracy parameters ("guaranteed error bounds") are used to control

the clustering; an accuracy parameter can be the percentage by which the number of

points is allowed to deviate from an expected value or the probability of a tile satisfying the accuracy criterion (column 9, line 63 to column 10, line 42); Table 1 shows age, salary, and years employed as "different columns" of a data table (column 5, lines 25 to 39);

"a model selector, associated with said table modeler, that selects a subset of said at least one model to form a basis upon which to compress said data table to form a compressed data table" – a data mining engine 12 forms conclusions about the accuracy of an initial model (M), and the model is refined until the model more accurately represents the data stored in the database (column 9, lines 37 to 62); a cluster must satisfy an accuracy requirement for the model to be judged suitable (column 10, lines 33 to 42); a model represents a compressed version of records in data database 10 (Abstract); a model is formed by selecting "a subset of said at least one model" at least because outlier data points, which have distances greater than a constant ξ for a cluster, are not members of clusters if a specified memory condition is exceeded (column 18, line 25 to column 19, line 13).

Claim Rejections - 35 USC § 103

5.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6.   Claims 2, 4, 5, 7, 8, 10, 12, 13, 16, 18, 20, 21, and 24 are rejected under 35

U.S.C. 103(a) as being unpatentable over *Fayyad et al. ('882)* in view of *Agrawal ('311)*.

Concerning claims 2, 10, and 18, *Fayyad et al. ('882)* does not disclose specifics

about the modeling process as employing classification and regression tree (CaRT)

data mining to model attributes. However, *Agrawal ('311)* suggests data mining with

decision trees for modeling records having one or more attribute values may be by

classification and regression trees. (Column 5, Line 63 to Column 6, Line 7; Column 6,

Lines 58 to 67) The stated objective is provide an efficient method for generating a

decision-tree classifier that is compact, accurate, has short training times, and is

scalable. (Column 3, Lines 11 to 24) It would have been obvious to one having

ordinary skill in the art to employ classification and regression trees for data mining of

model attributes as taught by *Agrawal ('311)* in the multi-dimensional database record

compression of *Fayyad et al. ('882)* for the purpose of generating decision trees by a

classifier that is compact, accurate, has short training times, and is scalable.

Concerning claims 4, 12, and 20, *Agrawal ('311)* discloses pruning for short

training time (column 8, line 40 ff).

Concerning claims 5, 13, and 21, *Agrawal ('311)* discloses pruning for

representing misclassification errors based upon encoding costs (column 9, lines 34 to

54), which is equivalent to a "scoring-based method".

Concerning claim 7, *Agrawal ('311)* discloses data mining with decision trees for

modeling records having one or more attribute values may be by classification and

regression trees (column 5, line 63 to column 6, line 7; column 6, lines 58 to 67);

implicitly, values of attributes are stored as models and not as data points, so values

"are not explicitly stored therein."

Concerning claims 8, 16, and 24, *Agrawal ('311)* discloses a greedy algorithm

may be used for subsetting (column 8, line 3).

7.   Claims 2, 3, 10, 11, 18, and 19 are rejected under 35 U.S.C. 103(a) as being

unpatentable over *Fayyad et al. ('882)* in view of *Pednault*.

Fayyad et al. ('882) does not disclose specifics about the modeling process as

employing classification and regression trees or a Bayesian network. However,

Pednault teaches a method for constructing predictive models that involve Bayesian

networks (column 2, lines 20 to 30 and column 2, lines 45 to 52) and classification and

regression trees (column 2, lines 35 to 45). The objective is to provide a method of

handling missing values. It would have been obvious to one having ordinary skill in the

art to employ classification and regression trees or Bayesian networks as suggested by

Pednault in the multi-dimensional database record compression of *Fayyad et al. ('882)*

for the purpose of providing a method for handling missing values.

8.   Claims 6, 14, and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable

over *Fayyad et al. ('882)* in view of *Chakrabarti et al. ('005)*.

Fayyad et al. ('882) omits selecting a subset based upon a compression ratio.

However, *Chakrabarti et al. ('005)* teaches a method for data mining where a

compression ratio is an indicator of complexity of compressed files. (Column 16, Lines

18 to 25) The objective is to select candidate data patterns from a dataset based on the variations of support values of a pattern. (Column 5, Lines 4 to 14) It would have been obvious to one having ordinary skill in the art to select a data subset based upon a compression ratio as suggested by *Chakrabarti et al.* *('005)* in the multi-dimensional database record compression of *Fayyad et al.* *('882)* for the purpose of selecting candidate data patterns from a dataset.

9.   Claim 15 is rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* *('882)* in view of *Agrawal et al.* *('048)*.

Fayyad et al. *('882)* does not disclose that a process by which a model selector selects a subset is NP-hard. However, *Agrawal et al.* *('048)* teaches that, in general, an optimized rule mining problem is NP-hard. (Column 4, Lines 9 to 14) The objective is to provide a method for identifying database association rules which are optimal at upper and lower support-confidence borders. (Column 4, Line 30 to Column 5, Line 45) It would have been obvious to one having ordinary skill in the art that model selection is an NP-hard algorithm as suggested by *Agrawal et al.* *('048)* in the multi-dimensional database record compression of *Fayyad et al.* *('882)* for the purpose of providing optimal association rules at upper and lower support-confidence borders.

Allowable Subject Matter

10.   Claim 23 is objected to as being dependent upon a rejected base claim, but

would be allowable if rewritten in independent form including all of the limitations of the

base claim and any intervening claims.

Response to Arguments

11.   Applicants' arguments filed 25 August 2005 have been fully considered but they

are not persuasive.

Firstly, Applicants argue that a clustering model is not an attribute in a data table

in terms of other attributes in the data table. Applicants say that *Fayyad et al. ('882)*

discloses a clustering model based on a mean for each dimension of data within a

database. Applicants recognize that *Fayyad et al. ('882)* discloses attributes involving

"years employed" versus "salary" in Table 1, but contend that there is no model of an

attribute using clusters. Thus, Applicants maintain that *Fayyad et al. ('882)* does not

model attributes in terms of other attributes but instead uses clusters determined from a

plot of dimensions of the database to model the database. This position is traversed.

A clustering model does present information about attributes in terms of other

attributes. *Fayyad et al. ('882)* clearly shows modeling one attribute in terms of another

attribute by a clustering model in Figure 5. A graph is a two-dimensional representation

of a relationship between two attributes. More generally, an n-dimensional space

represents relationships between n attributes, where a two-dimensional graph can be

thought of as a plane through an n-dimensional space. Figure 5 shows an x-axis of a

two-dimensional graph as "salary" and a y-axis of a two-dimensional graph as "years

employed". A plurality of data points on the graph represent a relationship between

"salary" and "years employed". Figure 6 shows how the data points on the graph of

Figure 5 are reduced to clusters in a clustering model. Cluster 1 is represented by a

Gaussian G1 having Mean $X(bar)^1$ and Cluster 2 is represented by Gaussian G2 with

Mean $X(bar)^2$. (Column 6, Lines 42 to 51) One skilled in the art can readily see how

the data points are grouped into clusters in Figure 5. Generally, Figure 5 clearly shows

a relationship between "salary" and "years employed" by a clustering model indicating

that salary increases with years employed. Specifically, Figure 5 shows a given value

of "salary" translates into a given value of "years employed". If a value of "salary" were

20, then the clustering model would indicate a value of "years employed" as, *e.g.,* 14.3.

If a value of "salary" were 40, then the clustering model would indicate a value of "years

employed" as, *e.g.,* 16.9. A clustering model reduces a set of data points to clusters,

but retains information relating one attribute in terms of other attributes. Thus, a

clustering model is advantageous for data mining because information about one

attribute in terms of another attribute can be readily determined by statistical

techniques.

Secondly, Applicants argue that *Fayyad et al. ('882)* does not teach selecting a

subset of the at least one model of the at least one attribute to form a basis upon which

to compress the data table to form a compressed data table. Applicants say that

Fayyad et al. ('882) instead selects clusters to model a database. Applicants note that

Fayyad et al. ('882) discloses selected clusters may be fine tuned to provide a better

model, but that this is not equivalent to forming a basis upon which to compress. This

position is traversed.

Fayyad et al. ('882) at least discloses the limitation of selecting "a subset of said

at least one model to form a basis upon which to compress" by excluding outlier points.

In at least some instances, outlier data points are excluded from clusters. Fayyad et al.

('882) discloses a data point is defined as an outlier if a distance of one dimension x_i of

a data point from cluster mean μ_i exceeds a constant ξ. If a specified memory limitation

is exceeded, then these outliers are not stored and an outlier is not taken into account

for a cluster by its Gaussian. (Column 18, Line 25 to Column 19, Line 13) At least for

this reason, Fayyad et al. ('882) discloses selecting "a subset of said at least one model

to form a basis upon which to compress" by excluding data points that are outliers for a

model when memory conditions are exceeded. If points of a model are excluded from a

model, then the model is a subset of a model.

Fayyad et al. ('882)'s method of selecting data for a model is equivalent to

Applicants' method of compressing a model, as disclosed by the Specification, Pages

10 to 12, ¶'s [0030] to [0034]. Here, it is disclosed that a table compression system

produces a compressed version of an input table T that selects a subset of data values

that are retained such that they predict the data within a prescribed degree of accuracy.

Equivalently, Fayyad et al. ('882) uses a plurality of cluster means as a subset of data

values that predict the data. Partitioning the data into more or fewer clusters, where

each cluster has a cluster mean, according to an accuracy criterion iteratively refines an

accuracy of a model. (Column 9, Line 63 to Column 11, Line 27) A mean of a cluster is

a data point corresponding to Applicants' retained data point. Thus, the number of

clusters of *Fayyad et al. ('882)* corresponds to the number of points retained by

Applicants.

It is true that *Fayyad et al. ('882)* discloses clusters to compress a model, while

Applicants use CaRTs, which are decision trees. However, independent claims 1, 9,

and 17 do not expressly disclose CaRT. Although the claims are interpreted in light of

the specification, limitations from the specification are not read into the claims. See *In*

re Van Geuns, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). Moreover, both

clusters and decision trees involve placing data points into classes.

Therefore, the rejections of claims 1, 9, and 17 under 35 U.S.C. §102(e) as being

anticipated by *Fayyad et al. ('882)*, of claims 2, 4, 5, 7, 8, 10, 12, 13, 16, 18, 20, 21, and

24 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al. ('882)* in view of

Agrawal ('311), of claims 2, 3, 10, 11, 18, and 19 under 35 U.S.C. §103(a) as being

unpatentable over *Fayyad et al. ('882)* in view of *Pednault*, of claims 6, 14, and 22 under

35 U.S.C. §103(a) as being unpatentable over *Fayyad et al. ('882)* in view of

Chakrabarti et al. ('005), and of claim 15 under 35 U.S.C. §103(a) as being

unpatentable over *Fayyad et al. ('882)* in view of *Agrawal et al. ('048)*, are proper.

Conclusion

Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Martin Lerner whose telephone number is (571) 272-
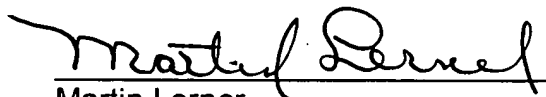
7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to

Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone

number for the organization where this application or proceeding is assigned is 703-

872-9306.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free).

ML
9/15/05

Martin Lerner
Examiner
Group Art Unit 2654